

Northumbria Research Link

Citation: Elharrouss, Omar, Almaadeed, Noor, Al-Maadeed, Somaya and Bouridane, Ahmed (2021) Gait recognition for person re-identification. The Journal of Supercomputing, 77 (4). pp. 3653-3672. ISSN 0920-8542

Published by: Springer

URL: <https://doi.org/10.1007/s11227-020-03409-5> <<https://doi.org/10.1007/s11227-020-03409-5>>

This version was downloaded from Northumbria Research Link:
<http://nrl.northumbria.ac.uk/id/eprint/46110/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)



**Northumbria
University**
NEWCASTLE



UniversityLibrary



Gait recognition for person re-identification

Omar Elharrouss¹ · Noor Almaadeed¹ · Somaya Al-Maadeed¹ ·
Ahmed Bouridane²

Published online: 27 August 2020
© The Author(s) 2020

Abstract

Person re-identification across multiple cameras is an essential task in computer vision applications, particularly tracking the same person in different scenes. Gait recognition, which is the recognition based on the walking style, is mostly used for this purpose due to that human gait has unique characteristics that allow recognizing a person from a distance. However, human recognition via gait technique could be limited with the position of captured images or videos. Hence, this paper proposes a gait recognition approach for person re-identification. The proposed approach starts with estimating the angle of the gait first, and this is then followed with the recognition process, which is performed using convolutional neural networks. Herein, multitask convolutional neural network models and extracted gait energy images (GEIs) are used to estimate the angle and recognize the gait. GEIs are extracted by first detecting the moving objects, using background subtraction techniques. Training and testing phases are applied to the following three recognized datasets: CASIA-(B), OU-ISIR, and OU-MVLP. The proposed method is evaluated for background modeling using the Scene Background Modeling and Initialization (SBI) dataset. The proposed gait recognition method showed an accuracy of more than 98% for almost all datasets. Results of the proposed approach showed higher accuracy compared to obtained results of other methods result for CASIA-(B) and OU-MVLP and form the best results for the OU-ISIR dataset.

Keywords Gait recognition · Angle estimation · Motion detection · Convolutional neural networks

✉ Omar Elharrouss
elharrouss.omar@gmail.com

Extended author information available on the last page of the article

1 Introduction

Across multiple cameras, person recognition and identification are important targets for many computer vision applications, especially monitoring systems [1]. The operation of recognition of a person from a set of images captured by several cameras is called person re-identification. The similarity measures can be the key to compute the matching between two or a set of images. However, the re-identification using video clips can be a problem for many applications [2], for example people tracking across multiple cameras. The video sequences captured by different cameras should be analyzed to re-identify the person and keep tracking him across all cameras in the surveilled areas.

The sequential methods that use a list of features are not efficient for the re-identification of persons due to several limitations such as differences between the analyzed objects in terms of shape, colors, scales, and others [3]. which, in turn, implies that the use of a limited number of features cannot be enough for proper identification. On the other hand, with deep learning techniques, the use of different and non-limited features for learning becomes a good alternative for solving person re-identifications problems. For that, to train these methods, a big scale of data is required from multiple camera views [4]. In addition, using preprocessing techniques the performance can help the learning model to better learning.

Owing to the differences between multiple images of the same person captured by different cameras, the re-identification can be difficult even with deep learning models. The difference can be in the shape, the color of the cloths, as well as the scale of the images. Contrarily, the gait is a feature that cannot be changed and is repetitively performed by the person; thus, it could be considered as an identification feature [5, 6]. The gait of each person can be used for re-identifying a person across different cameras in Fig. 1.

Human gait is generally observed to be a uniquely human characteristic that is difficult to replicate or hide. Hence, it represents a critical biometric identification feature for human identification [7]. Therefore, gait identification has been recognized as an essential identification technology for different applications in crime control and detection systems in high security, civilian, and public areas such as airports, stations, banks, and military bases [8, 9]. Recently, gait recognition has also found significant application in gender, age, and ethnicity prediction systems and in cyber-physical healthcare systems enabled through connected wearable devices [10–13].

Person re-identification using human gait can be an efficient identification technique that can overcome the person re-identification problems including shape, colors, and scales. The gait analysis can be a good solution due to its unicity for each person; nonetheless, some limitations can affect the performance of such a gait recognition algorithm. One of these limitations is the view angle of each gait because there is a remarkable difference between the gait of a person. To handle the gait recognition under different view angles, we propose a multitask-based method using convolutional neural networks (CNNs) on gait energy image (GEI) features. The proposed method starts by extracting the gait energy image (GEI) of



Fig. 1 Person re-identification across cameras

each person in a scene using the motion detection and segmentation of each person based on a proposed background subtraction method. Before recognizing the gait, the view angle is estimated using a proposed CNN model. Then, the verification of the estimated angle is performed using another proposed CNN model on GEIs. This technique aims to improve gait recognition accuracy. We train and test our proposed gait recognition method on three publicly available datasets. The proposed method achieves better results compared with some existing methods.

The rest of the paper is organized as follows: The literature overview related to our work is presented in Sect. 2. The proposed system is presented in Sect. 3. Experimental analysis is provided in Sect. 4. The conclusion and future works are given in Sect. 5.

2 Literature review

Gait recognition involves four main steps: image acquisition, preprocessing to segment out the binarized silhouettes from the background, training and/or feature extraction from silhouettes, and, lastly, classification or recognition of gait sequences by matching the testing and training feature space. Assuming the silhouettes to be obtained from fixed cameras observing static scenes, the silhouettes can be preprocessed by simple and less computationally expensive techniques like background segmentation [14]. After the preprocessing of the acquired images, features are extracted from the foreground silhouettes and several recognition techniques are used to match the feature space of the training sequences or “gallery” and test sequences or “probe.” Since the size of the feature vector involved in this scheme

becomes very large, the dimensionality of the feature vector may be reduced before classification/recognition. For this purpose, principal component analysis (PCA) and multiple discriminant analysis (MDA) are used to achieve good reduced data representation by discarding the features which show low variance.

Since human gait is a human behavior that involves movement of various parts of a human's body, the early research on gait recognition started with attempts to model the human body as a whole. Hence, this approach was given the name of the model-based approach to gait recognition. In the model-based approach, after acquiring the images, silhouettes are obtained by noise removal and binarization of the walking person's 2D image. From the silhouettes, specific parameters of the human model are extracted. The parameters typically pertain to various body parts—for example, lengths of body parts (typically limbs); widths of body parts like head, torso, knees, and arms; the position of head and shoulders; and trajectory defined by hip joint angles. In fact, the hip rotation pattern from the sequence of images carries essential information for gait recognition. The authors in [15] defined 22 such parameters of human body parts involved in the gait and used these to form a layered deformable model of the human body. The parameters of the human body model can be updated over time for the individual silhouettes obtained from the sequential motion frames and hence used to represent the gait. Although the model-based approach has shown the ability to obtain good gait recognition, yet, the recognition has faced some challenges. For instance, the images have many occlusions and shadows, and locating the body segments from the binarized image silhouettes is difficult. For increased recognition accuracy, modeling the human body often calls for the conversion of excellent quality 2D images to 3D computer models, which is a complicated and compute-intensive task [7]. Moreover, the quality of images captured by surveillance cameras is quite weak which adversely affects the quality of gait recognition. Therefore, the focus of further research shifted more toward the model-free approach.

Model-free, also called motion-based approach, can again be categorized into two types: sequential motion-based and spatiotemporal motion-based approaches. In sequential motion approaches, gait is represented as a time sequence of human poses, while the spatiotemporal approach represents gait by mapping the distribution of motion through space and time [14]. The sequential motion-based approach proposed in [34] involves representing the motion through temporal templates that identify where the motion has occurred and also recording the history of these motions. The spatiotemporal approaches proposed in the literature differ primarily in the preprocessing, feature extraction, and classification techniques used for the silhouette-based gait sequences. The authors in [35] proposed a feature selection mechanism called gait energy image (GEI) through which a history of gait movements is recorded in a single 2D template instead of storing it as a sequence of templates. The spatiotemporal GEI feature is obtained by averaging the pixels of the silhouette in different frames of a gait cycle. Recognition involved statistical gait feature fusion of real and synthesized (distorted) gait templates. This approach does not only preserve space but also report high recognition performance. The authors in [9] proposed gait entropy image to be used as an automatic feature selection mechanism for the gallery (ground truth) and probe (testing) images. This feature selection scheme is proven to mitigate the effects of covariate walking conditions. The

proposed recognition approach, which is called the adaptive component and discriminant analysis (ACDA), is a fast recognition approach to gait recognition.

In the same context, many researchers have conducted studies and proposed different approaches that aim to handle recognition under different angles of view. The authors in [16] and [17] proposed a walking speed-invariant gait recognition method based on RBF neural networks. Other authors in [18] proposed a gait recognition method by extracting a couple of joint angles from two signature and then using a baseline 1-NN classifier to classify the gait. In the same context, Rida et al. [19] proposed a gait recognition approach based on the phase-only correlation. In [20], using convolutional neural networks, the authors proposed a gait recognition approach. For cross-view gait recognition, authors in [23] proposed an approach based on a model named video transformation model (VTM). For that same purpose, Wu et al. [21] proposed a method using the CNN model. Using tensor representation, the authors in [36] proposed an approach for cross-view gait recognition. Spatiotemporal HOG features are used also for cross-view gait recognition in [25].

Recently, many methods have been proposed for handling the angle variation exploiting the silhouette sequences [26, 28] or proposing some other features [22, 32], or using GEI images for training their deep learning models [27]. Authors in [26] proposed a gait recognition method named localized Grassmann mean representatives with partial least squares regression (LoGPLS), whereas authors in [22] proposed new features called autocorrelation feature where the image at lag time zero is similar to GEI. Another new feature is proposed in [32] and called period energy image (PEI) which is a multi-channel gait template. The proposed feature is used for gait recognition. In addition, the authors proposed a gait recognition method based on the local GEI (LGEI) feature with a self-adaptive hidden Markov model (SAHMM) [29]. Instead of using GEI or other similar features, some authors trained their models on silhouette sequences, like in [28]. In [30], the authors used the silhouette sequence to train a deep learning method based on ResNet and LSTM. In order to recognize the gaits on a very large scale of data, generative adversarial networks (GAN) are used recently in many works [22, 27, 37]. The recognition in [27] is performed on 10000 subjects, whereas in [37] two-stream GAN model is used to learn gait features. Authors in [32] have used two gait templates: GEI and PEI with a GAN model for recognizing gaits under different angles of view. In the same context, authors in [31] used RGB image sequences as an input of the proposed architecture based on autoencoder networks and LSTM. Using silhouette-based features captured by specified cameras, the authors in [33] proposed a CNN model for predicting the angle and also used it for recognizing the gait.

The obtained results by different methods for gait recognition are convincing, but there is still a chance for improved efficiency. Table 1 represents a summarization of the cited gait recognition methods. The deep learning methods using CNNs or GAN improve the performance of the recognition, but the complexity of each new dataset disables existing models to handle the new challenges like the variation of the angle of view, such as the dataset in [38] representing 14 angles of view where [39] represented just 11. So for each new dataset, the existing method cannot be suitable to recognize the gait on it.

Table 1 Summarization of gait recognition methods

Method	Feature	Technique
[16, 17]	Walking speed	Radial basis function (RBF) neural networks
[18]	Motion capture data	1-NN classifier
[19]	Gait energy image	Phase-only correlation (POC)
[20–22]	Gait energy image	Convolutional neural network (CNN)
[23, 24]	Gait energy image	Video transformation model (VTM)
[25]	HOGs	Two-sided Fourier series
[26]	Partial least squares regression (LoGPLS)	Localized Grassmann mean representatives
[27]	Gait energy image	Generative adversarial networks (GAN)
[28]	Gait sequence	Convolutional neural network (CNN)
[29]	Local gait energy image (LGEI)	self-adaptive hidden Markov model (SAHMM)
[30, 31]	Gait sequence	CNN + long short-term memory (LSTM)
[32]	Gait sequence, PEI	Generative adversarial networks (GAN)
[33]	Joints relationship pyramid mapping (JRPm)	Convolutional neural network (CNN)

3 Proposed approach

Gait recognition technology is an efficient technique to re-identify a person during a pass across multiple cameras. This is due to that gait represents an effective measure for person identification from distance owing that it is a unique characteristic of each person, and unlike the other recognition measures, gait cannot be different in multiple images of the same person as the case for shape, clothes colors, and scale that might vary from an image to another. However, the change in view angles during visual surveillance scenes is a common challenge that often faces the gait recognition. Therefore, in this work, we attempt to deal with this challenge. First, each person's silhouette, after the background subtraction-based methods for motion detection, extraction method is proposed.

After the extraction of the binary sequence of each target, the GEI of each sequence is extracted as illustrated in Fig. 2. To recognize the person's gait, the proposed approach combines the estimation of the view angle before starting the recognition using a multitask CNN model. As illustrated in Fig. 2, the view angle falls within a range of [0–270] similar to the example presented in Fig. 3 obtained from the OU-MVLP dataset [38]. Here, the angle is estimated from the GEI image using a CNN model. Then, the recognition is performed using the second CNN model. A detailed description of each step is represented in this section. For the training, a set of datasets is used including CASIA-(B) [24], OU-ISIR [39], and OU-MVLP [38] datasets.

3.1 People detection and tracking

To ensure accurate detection of the moving human's body silhouette in the scene, a background subtraction-based approach is proposed. The method starts with

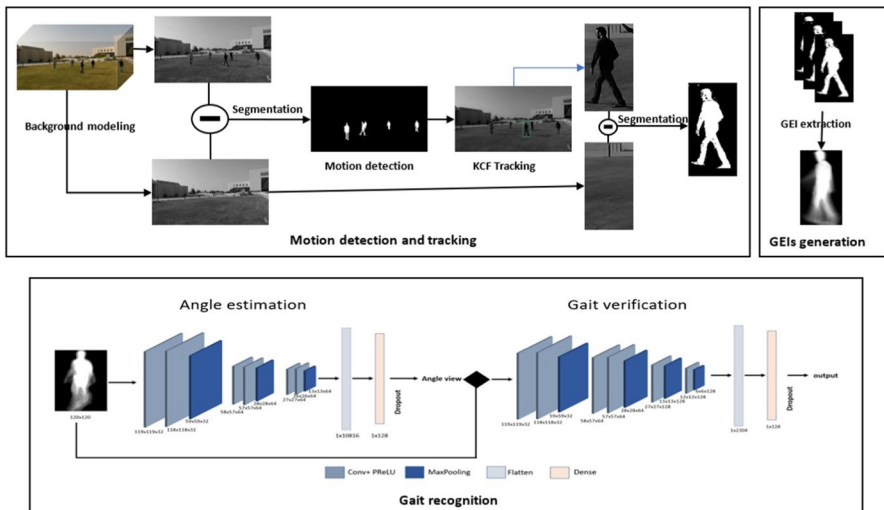


Fig. 2 Flowchart of the proposed method

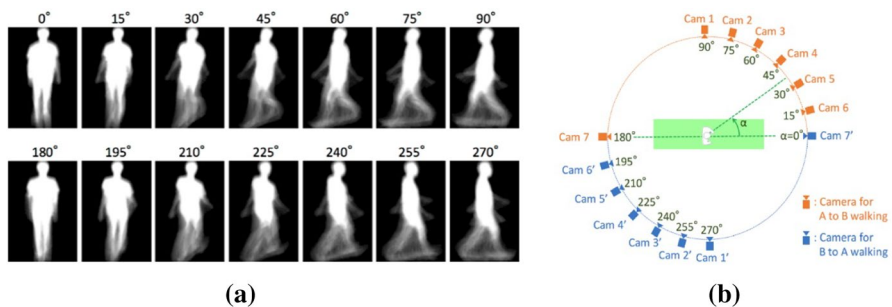


Fig. 3 Examples from OU-MVLP dataset. **a** GEI for each view angle. **b** Cameras distribution

background modeling, in which the main step background subtraction-based method consists of extracting the unchanged pixels and regions in an image sequence [40].

The modeling starts by dividing each frame into $w \times w$ blocks and then computing the similarity between blocks $b(i,j)$ of consecutive images using Equation (1) of algorithm 1. The similarity is computed using Equation (2) defined in [41]. The background model is generated by collecting the maximum values of the sum of similarity of each block (i,j) ; regions of the blocks that did not change a lot during the 100 frames will have the most significant values because the value is 1 where two blocks are similar. The generated background model is defined based on SS values using Equation (3).

The following step is the background subtraction, where the background is subtracted from each current frame of the video using the absolute difference. Then, based on the subtraction results, a segmentation operation is performed to classify the pixels belonging to the background and those belonging to the foreground or

the moving objects. Here, an adaptive threshold is used where the method tests a set of thresholds and makes the selection of the one that gives the best results. In this paper, we propose a segmentation method for selecting this threshold adaptively using the exponential function of the absolute difference between the current frame and background frame. The equation is expressed in Equation (4) in algorithm 1. Here, the values of T have to be within the range of $[0,1]$, I_t is the current frame and B_t denotes the background image. The threshold value converges to 0 when the background subtraction result goes to 0, and the threshold values tend to 1 when the background subtraction value is significant.

The computation of moving objects at each time in the video which are represented by a binary image is performed using the selected threshold. The binary frame at time t of the video is computed using Equation (5) in algorithm 1. After the generation of the binary image that represents the detected moving objects, an update of the background model is performed.

Algorithm 1 Motion detection

Input: Image sequence,

Output: Background model B ,

1: **Similarity computation:**

for $i \leftarrow 1, 2, \dots, n$; $j \leftarrow 1, 2, \dots, m$ do

$$SS_{b(i,j)} \leftarrow \sum_{t=1}^{T-1} \sum_{i=1}^w \sum_{j=1}^w \cosin(I_t^{(i,j)}, I_{t+1}^{(i,j)}) \quad (1)$$

$$\cosine(a, b) = \frac{\sum_i^n a_i b_i}{\sum_i^n a_i^2 \sum_i^n b_i^2} \quad (2)$$

end for

2: **Background extraction:**

for $i \leftarrow 1, 2, \dots, n$; $j \leftarrow 1, 2, \dots, m$ do

$$B^{(i,j)} \leftarrow \text{Argmax}\{SS_{(i,j)}\} \quad (3)$$

end for $B \leftarrow B^{(1,1)}, B^{(1,2)}, B^{(i,j)}, \dots, B^{(n,m)}$

3: **Background subtraction and threshold selection:**

for $i \leftarrow 1, 2, \dots, n$; $j \leftarrow 1, 2, \dots, m$ do

$$T(i, j) \leftarrow 1 - e^{-|I(i,j) - B(i,j)|} \quad (4)$$

end for

4: **Foreground segmentation:**

for $i \leftarrow 1, 2, \dots, n$; $j \leftarrow 1, 2, \dots, m$ do

$$D(i, j) \leftarrow \begin{cases} 255 & \text{if } T(i, j) \simeq 1 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

end for

3.2 Gait recognition

Most of the gait recognition methods employ the GEI feature or image sequence of the detected silhouette. For GEI-based methods, they start by extracting moving human silhouettes from the video using a Gaussian mixture model or background subtraction method like we do. Then, GEI is computed by averaging the silhouettes during the used sequence. An example of GEIs extracted and the angle of view of each one from the OU-MVLP dataset is represented in figures 3.

After the extraction of GEIs for each person from the videos (some datasets give GEI), the gait recognition process is performed. In this paper, we proposed a multitask CNN model for gait recognition by estimating the angle of views for each gait before starting the recognition. Figure 4 represents the procedure of GEI extraction and the process of estimating the angle. The following section discusses the proposed multitask architecture for gait recognition.

3.2.1 Angle estimation model

Gait images captured from different view angles can affect the gait recognition accuracy of any method as it is difficult for a system to estimate and recognize an identity from GEI images under different angles. The degree of gait variation has been proposed in many datasets that were captured using multiple cameras from different angles. Many works have been proposed to recognize the gait under this challenge but without estimating this angle. For example, all methods recognize the gait on each angle directory that is specified on the datasets. However, when we have a novel gait, we need to recognize the angle of capturing, and this was not handled by most of the previously proposed methods. To handle this problem, we proposed a multitask gait recognition method, using two collaborative CNN models. The first model aims to recognize the angle of capturing, and the second is to recognize this gait.

The proposed CNN model, as shown in Fig. 5, of angle estimation, is trained on data from the three datasets including CASIA-(B) [24], OU-ISIR [39], and OU-MVLP [38]. The number of subjects in the three datasets is about 14K where each subject is under several angles of view. As a first step, each dataset has been trained on a variation of capturing angles. The accuracy of angle estimation using our proposed model reaches 98%.

3.2.2 Recognition model

The selection of optimal CNN architecture is a challenging problem that depends on the application. In this paper, a multitask CNN, which is supervised learning with a multistage deep learning network, has been implemented. Multitask CNN could learn multiple stages of invariant features from the input images. Convolution and pooling layers are the main layers in a CNN model. Any involved CNN

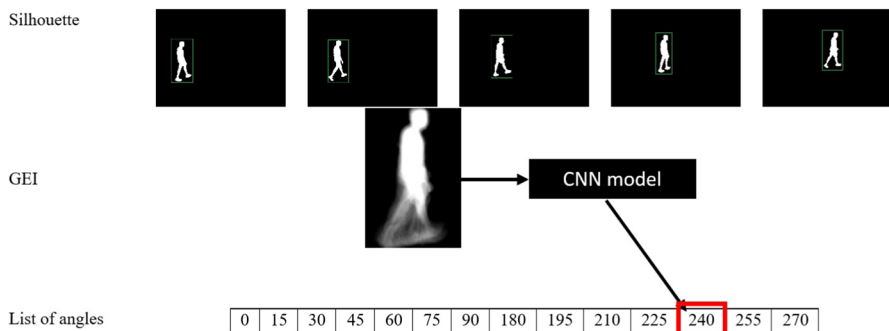


Fig. 4 GEI extraction and angle estimation process

can be constructed with a couple of combination convolution–pooling. Hence, using images as inputs and backpropagating the errors, the learning takes place.

The architecture of the proposed model, as illustrated in Fig. 6, composes of 2–3 convolution–pooling units, with three convolutional layers and three Max-Pooling layers, one flattened layer, and two fully connected layers. The output layers comprise of ten neurons that represent the number of actions. Each convolution neural network is referred to in this work as the following: $I(x,y,f)$ as an input image with a size of $x \times y$ and f number of channels; $\text{Conv}(x,y,k)$ is the convolutional layer and pooling $\text{Mpool}(x,y,k)$ where x and y are image dimension, f number of channels, and k number of kernels. PReLU indicates parametric rectified linear unit, $\text{FC}(n)$ is a fully connected layer with n neurons, and $D(r)$ is a dropout layer with a dropout ratio r .

As an activation function, we use the parametric rectified linear unit (PReLU), which is a generalized parametric formulation of ReLU. Using this activation function, the parameters of rectifiers are learned adaptively and improve the accuracy with a negligible extra computational cost [42]. Only positive values are fed to the ReLU activation function, while all negative values are set to zero. PReLU

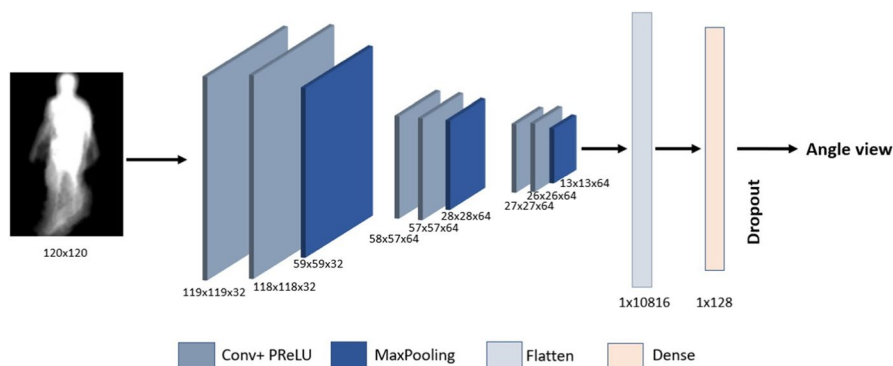


Fig. 5 Angle estimation CNN model

assumes that a penalty should be applied for negative values, and it should be parametric. The PReLU function can be defined as:

$$f(y_i) = \begin{cases} y_i & \text{if } y_i > 0 \\ a_i y_i & \text{if } y_i \leq 0 \end{cases}$$

where a_i controls the slope of the negative part. When $a_i = 0$, it operates as ReLU, and when a_i is a learnable parameter, it is referred to as parametric ReLU (PReLU). Figure 5 shows the shape of PReLU activation. If a_i is a small fixed value, PReLU becomes LReLU ($a_i = 0.01$). PReLU can be trained using the backpropagation concept.

The input of the system is an image of gait with a resolution of 120 x 120 pixels. The model is trained using CrossEntropy with a batch size of 20 examples and a learning rate of 0.001 as described in Table 2.

4 Experimental results

In this work, we present a gait recognition method for person re-identification. The proposed method contains two phases: first the detection of moving objects based on background subtraction and then the gait recognition which is performed by estimating the angle of view and then recognizing the gait. The proposed method for background modeling has been tested on SBI dataset and then compared with some existing methods. Our approach involves the recognition of gait images based on viewing angle estimation of the GEIs of input probe images. GEI is the temporal average of the gait silhouettes in a gait cycle. Angle estimation is performed using our proposed CNN-based angle learning model. From the estimated angle, we recognize the gait of the subject by using the CNN classifier on the gallery GEIs for the corresponding angle. We evaluate our proposed method using three publicly available datasets, namely CASIA-(B) gait database;

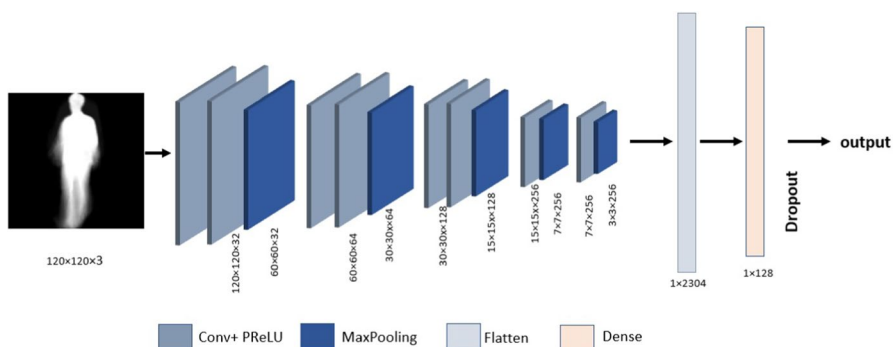


Fig. 6 Gait recognition CNN model within each angle

OU-ISIR large population gait database, and OU-MVLP multi-view large population database.

For the OU-ISIR and CASIA-(B) datasets, performance of our method was compared against several benchmark methods including [21, 23, 31–33, 36].

Since our method involves gait angle estimation prior to gait recognition, we needed datasets with gait images acquired from different viewing angles to evaluate the performance of our approach. We chose the databases of cross-view gait images because all these three datasets provide gait images of subjects from different viewing angles. Below, the gait datasets used to test our method are briefly described.

4.1 Datasets

CASIA-(B) Dataset CASIA-(B) dataset [24] provides gait data of 124 subjects, captured from 11 different viewing angles from 0 to 180 degrees. The angles are equally spaced at intervals of 18 degrees. In addition to the variation in view angle, the gait data are captured also for different clothing and carrying conditions for each subject. The data consist of videos and silhouettes extracted from video files. The CASIA-(B) dataset also provides images for each subject corresponding to different carrying conditions, e.g., bags and clothes; but this work is limited to the six images provided for normal walking conditions as it forms the major part of the gallery.

OU-ISIR Dataset OU-ISIR provides gait images of 4007 males and females from different ages (1–94) subjects captured by two cameras from four observation angles, i.e., 55, 65, 75, and 85 degrees. The observation angle is defined as the y-axis of the line of sight of camera in the world coordinate system (parallel to walking direction) [39]. A bin is created for each of these angles for camera A and B, and a subject recorded in a particular angle by a camera is placed in the corresponding bin of that camera. Size-normalized silhouettes, or the GEI features, are provided in the dataset for each subject.

OU-MVLP Dataset The OU-MVLP dataset tries to overcome the problem of overfitting because of the small sample size by providing a large number of gait images of 10,307 subjects of ages between 2 and 37 years, captured from 14 different view angles [38]. The angles range from 0 to 90 degrees when the subject is walking from point A to point B and 180 to 270 degrees for the opposite direction. Seven cameras are fixed at 15 degree intervals in the ranges mentioned above, so 28 images can be recorded for each subject. It is the largest known dataset so far, and to the best of our knowledge, it has been never evaluated an approach on OU-MVLP.

Table 2 Training hyper-parameters (general classifier)

Optimizer	LR	Epsilon	Beta_1	Beta_2	Decay	Epochs	Batch size
Adam	0.001	1e-08	0.91	0.999	0	50	20

4.2 Background modeling evaluation

SBI dataset is used to evaluate the proposed method for background modeling. Figure 7 represents the generated background using the proposed approach. The obtained results are convincing, where using our method the background is built without artificial ghosts for all videos. For “Foliage” and “People & Foliage” sequences, the proposed method could successfully estimate the background with good results even the sequences a full of moving objects during all time of videos.

In order to consolidate the visualized results, we use different metrics, including gray-level error (AGE), total number of error pixels (EPs), percentage of error pixels (pEPs), total number of clustered error pixels (CEPs), peak signal-to-noise ratio (PSNR), multiscale structural similarity index (MS-SSIM), color image quality measure (CQM).

These metrics are presented in Table 3 that illustrates the obtained results compared with two background modeling methods used in IMBS-MT [43] and [44], respectively. As shown, the proposed method succeeds to modelize the background with good results in comparison with other methods in the most dataset videos including HighwayI, Hall & Monitor, Snellen, and Foliage.

In addition to the cited metrics, precision and recall metrics are used for evaluating the proposed background modeling approach. Figure 8 illustrates precision and recall values for the generated background model for each image sequence.

4.3 Performance evaluations

To evaluate the angle estimation model, we trained our model on each dataset including CASIA-(B), OU-ISIR, and OU-MVLP datasets. Using the proposed method, the accuracy rate for angle estimation reached 99% for CASIA-(B) and 98% for OU-ISIR and OU-MVLP dataset as illustrated in Table 4.

Evaluation on CASIA-(B) dataset

The evaluation of the effectiveness of the proposed gait recognition method on CASIA-(B) dataset is performed. The accuracy of the recognition with optimal parameters is reported. The comparison is made with three existing methods including [21, 23, 36]. The GEI feature is used for all compared methods for characterizing the gait patterns. For that, the recognition rate comparing with other methods results for GEIs under probe view 54° is illustrated in Table 5. From the table, it can be observed that the proposed method outperforms the other approaches, and for all angles, the average recognition rate reaches 97%. The recognition rate also achieves 98% for the angle in the interval $[36^\circ, 126^\circ]$. Comparing with [21], our obtained results are close. Also, for the opposite angles like 18° and 162° , the proposed approach reaches a similar recognition rate.

Table 3 Accuracy results of the compared methods on SBI dataset

Data	Method	AGE	pEPs%	pCEPS%	MSSSIM	PSNR	CQM
CaVignal	IMBS-MT [43]	0.7692	0.0147	0.0000	0.9982	45.9202	57.1044
	[44]	3.8855	0.0041	0.0000	0.9933	34.8725	54.5813
	Ours	1.1953	0.0287	0.0000	0.9971	43.6937	56.3661
Foliage	IMBS-MT [43]	7.5809	9.8507	3.1319	0.9090	22.7278	34.0028
	[44]	8.5594	0.4313	0.0000	0.9892	27.7099	39.6381
	Ours	1.8632	0.1277	0.0000	0.9972	36.9587	44.0911
Hall & Monitor	IMBS-MT [43]	1.5350	0.0923	0.0000	0.9954	38.6214	48.5224
	[44]	2.3878	0.1567	0.0102	0.9934	37.9820	61.3861
	Ours	1.1723	0.0855	0.0002	0.9980	40.6831	63.4881
HighwayI	IMBS-MT [43]	1.4913	0.0612	0.0026	0.9939	14.7728	58.8328
	[44]	3.0301	0.1855	0.0085	0.9880	35.0837	59.7762
	Ours	1.3602	0.0079	0.0000	0.9960	42.2736	62.7342
HighwayII	IMBS-MT [43]	1.8684	0.0260	0.0000	0.9960	40.1098	48.80094
	[44]	2.3279	0.1113	0.0000	0.9967	38.9867	49.7341
	Ours	1.9553	0.0111	0.0000	0.9947	38.6639	47.3772
People & foliage	IMBS-MT [43]	8.3982	7.3568	3.2305	0.8514	20.0658	32.5231
	[44]	5.7884	0.1974	0.0034	0.9885	35.7556	47.2501
	Ours	1.3903	0.0059	0.0000	0.9937	40.7648	47.6089
Snellen	IMBS-MT [43]	14.4480	25.3279	19.7290	0.8668	19.7436	40.115
	[44]	3.7620	0.0163	0.0000	0.9951	37.1563	49.3740
	Ours	1.6283	0.0202	0.1387	0.9976	37.7187	49.6327

The bold values represent the best results

Table 4 Accuracy results of the angle estimation model on different datasets

	CASIA-(B)	OU-ISIR	OU-MVLP
Accuracy(%)	99.1	98.7	98.4

**Fig. 7** Background results on SBI dataset using the proposed approach

Evaluation on OU-ISIR dataset

The accuracy evaluation of the proposed method is applied also on OU-ISIR dataset that contains two sequences for each subject. For that, the recognition rates

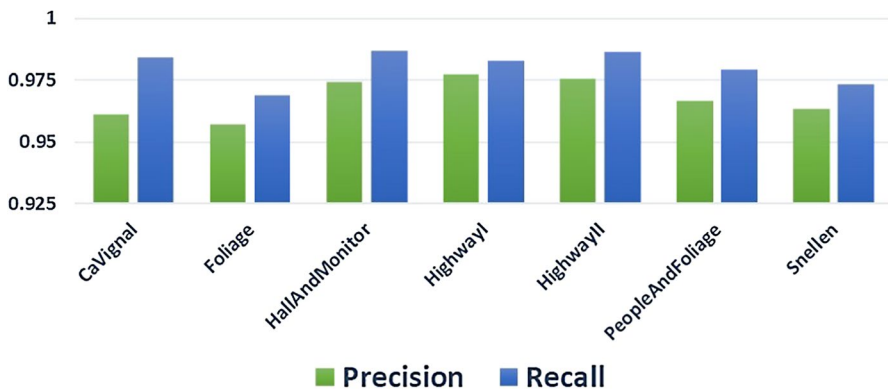


Fig. 8 Precision and recall values of the generated background models

are tested for each cross-view ten times after the estimation of the angle. Figure 8 illustrates the precision and recall values for the generated background model for each image sequence. As shown, the recognition rate decreases related to the difference between the probe and gallery views. Even the view difference is maxed like for 85° and 55°, the proposed method rates are stable and achieve 98%.

The proposed method is also compared with the recent methods available in the literature. The recognition rates are presented by the third digit after the decimal point because the rates are close. It can be observed that the proposed method can recognize the gait with a reasonable accuracy rate and close to the method in [36] that uses many features as inputs. Comparing with the other methods, our obtained results are better.

From the diagrams in Fig. 9 and the results in Table 5, results obtained by [21, 23], and the proposed method are stable, where the proposed method in [36] gives the best results on OU-ISIR dataset and the accuracy values can reach 100% for the probe views 65, 75, and 85, where the results are less than the proposed method and the method in [21] for CASIA-(B) dataset. Using the proposed method, the obtained accuracies are close and stable compared to the method in [36] and this is the reason

Table 5 Comparison on various methods on CASIA-(B)

Method	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	Average
Mu et al. [23]	0.21	0.67	0.96	–	0.97	0.7	0.66	0.39	0.33	0.20	0.22	0.531
Wu et al. [21]	0.88	0.95	0.98	0.96	0.94	0.92	0.94	0.97	0.97	0.96	0.86	0.939
He et al. [32]	0.63	0.73	0.79	0.81	0.75	0.71	0.73	0.80	0.80	0.77	0.63	0.74
Ben et al. [36]	0.43	0.78	0.99	–	0.98	0.82	0.77	0.76	0.57	0.42	0.35	0.687
Zhang et al. [31]	0.93	0.92	0.90	0.92	0.87	0.95	0.94	0.95	0.92	0.90	0.90	0.92
Liao et al. [33]	0.95	0.96	0.95	0.96	0.95	0.97	0.97	0.94	0.96	0.97	0.97	0.959
Ours	0.94	0.95	0.97	0.97	0.98	0.98	0.98	0.98	0.97	0.95	0.93	0.963

The bold values represent the best results

for dividing the recognition model in angle estimation model and recognition model for each angle.

Evaluation on OU-MVLP dataset

The proposed method is evaluated also on OU-MVLP dataset that provides gait images of 10,307 subjects captured from 14 different view angles [38]. The angles range from 0 to 90 degrees when the subject is walking from point A to point B, and 180 to 270 degrees when walking from point B to point A. Table 6 represents the recognition rates for all angles. As shown, the recognition achieves a rate of 98%.

In Fig. 10, the average rank 1 accuracies are reported on cross-view gait identification excluding the identical views as well as the representation of the accuracies of different gallery sizes on OU-MVLP dataset. It can be observed that some of the obtained results, including ones of the proposed method, are convincing even when it is performed on this large-scale cross-view gait recognition. The evaluations are made on 1800 identities for the method [22] and 1000 and 5000 for the proposed and [27] methods. The other method did not declare the number of subjects for these comparisons. The gallery sizes can differ, for example in indoor offices, and the gallery size is smaller than an outdoor scene. In Fig. 10b, we can see the performance tends to be higher with a gallery of 1000 identities. Also, we can observe that the accuracy decrease while the gallery size increases.

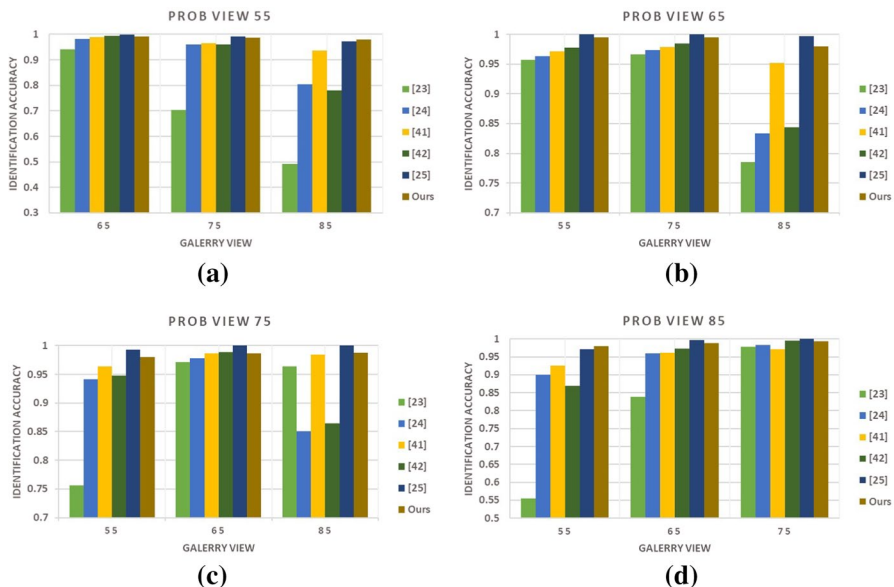


Fig. 9 Recognition rate of various method on OU-ISIR

Table 6 Comparison on various methods on OU-MVLP dataset

Method	0°	15°	30°	45°	60°	75°	90°	180°	195°	210°	225°	240°	255°	270°
[28]	0.79	0.87	0.89	0.90	0.88	0.88	0.87	0.81	0.86	0.89	0.89	0.872	0.87	0.86
[22]	0.79	0.89	0.93	0.95	0.95	0.95	0.95	0.86	0.90	0.95	0.95	0.93	0.94	0.94
Ours	0.93	0.95	0.95	0.97	0.98	0.97	0.98	0.92	0.94	0.95	0.95	0.97	0.97	0.98

The bold values represent the best results

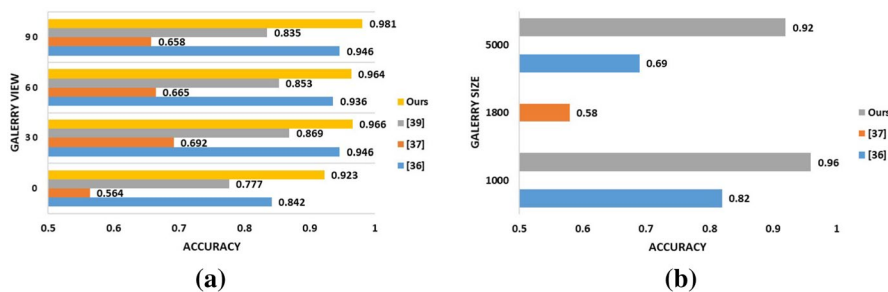


Fig. 10 Performance results on OU-MVLP dataset. **a** Rank 1 identification rates under probe 0°, 30°, 60°, 90° on OU-MVLP dataset. **b** Accuracy rates using different gallery sizes

5 Conclusions

Person re-identification is a challenging task for computer vision applications due to the variation of the appearances of the same person from different camera views. Cross-view of the gait is also posing a problem because different capturing angles limit the recognition of the gait. This paper presents a discriminant method to overcome this problem. Multitask method for gait recognition starts with the detection of people using the background subtraction method and then extraction of the GEIs for each person. After that, the proposed CNN-based model is used to estimate the angle before recognizing the gait. Experimental results exploited CASIA-(B), OU-ISIR, and OU-MVLP gait datasets which demonstrate that our multitask method is effective and on average more robust than other state-of-the-art methods.

Acknowledgements This publication was made by NPRP Grant # NPRP8-140-2-065 from Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

Funding Open Access funding provided by the Qatar National Library.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Jüngling K, Arens M (2011) View-invariant person re-identification with an implicit shape model. In: 2011 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 197–202
2. Liu Z, Zhang Z, Wu Q, Wang Y (2015) Enhancing person re-identification by integrating gait biometric. *Neurocomputing* 168:1144–1156
3. Gao B, Zeng M, Xu S, Sun F, Guo J (2016) Person re-identification with discriminatively trained viewpoint invariant orthogonal dictionaries. *Electron Lett* 52(23):1914–1916
4. Wu Y, Lin Y, Dong X, Yan Y, Ouyang W, Yang Y (2018) Exploit the unknown gradually: one-shot video-based person re-identification by stepwise learning. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5177–5186
5. Carley C, Ristani E, Tomasi C (2019) Person re-identification from gait using an autocorrelation network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*
6. Riachy C, Khelifi F, Bouridane A (2019) Video-based person re-identification using unsupervised tracklet matching. *IEEE Access* 7:20596–20606
7. Nambiar A, Bernardino A, Nascimento JC (2019) Gait-based person re-identification: a survey. *ACM Comput Surv (CSUR)* 52(2):33
8. Hossain E, Chetty G, Goecke R (2012) Multi-view multi-model gait based human identity recognition from surveillance videos. In: *IAPR Workshop on Multimodal Pattern Recognition of Social Signals in Human-Computer Interaction*, 88–99. Springer, Berlin, Heidelberg
9. Bashir K, Xiang T, Gong S (2010) Gait recognition without subject cooperation. *Pattern Recognit Lett* 31(13):2052–2060. <https://doi.org/10.1016/j.patrec.2010.05.027>
10. Sun F, Zang W, Gravina R, Fortino G, Li Y (2020) Gait-based identification for elderly users in wearable healthcare systems. *Inf. Fusion* 53:134–144. <https://doi.org/10.1016/j.inffus.2019.06.023>
11. Johnston AH, Weiss GM (2015) Smartwatch-based biometric gait recognition. In: 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems, BTAS 2015. <https://doi.org/10.1109/BTAS.2015.7358794>
12. Shila DM, Eyisi E (2018) Adversarial gait detection on mobile devices using recurrent neural networks. In: *Proceedings of 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications and 12th IEEE International Conference on Big Data Science and Engineering, Trustcom/BigDataSE 2018*, 316–321. <https://doi.org/10.1109/TrustCom/BigDataSE.2018.00055>
13. Zhao H, Wang Z, Qiu S, Wang J, Xu F, Wang Z, Shen Y (2019) Adaptive gait detection based on foot-mounted inertial sensors and multi-sensor fusion. *Inf Fusion* 52:157–166. <https://doi.org/10.1016/j.inffus.2019.03.002>
14. Balazia M, Plataniotis KN (2017) Human gait recognition from motion capture data in signature poses. *IET Biom* 6(2):129–137
15. Lu H, Plataniotis KN, Venetsanopoulos AN (2006) A layered deformable model for gait analysis. In: *FGR 2006 Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, 249–256. <https://doi.org/10.1109/FGR.2006.11>
16. Zeng W, Wang C (2015) Gait recognition across different walking speeds via deterministic learning. *Neurocomputing* 152:139–150. <https://doi.org/10.1016/j.neucom.2014.10.079>
17. Zeng W, Wang C (2016) View-invariant gait recognition via deterministic learning. *Neurocomputing* 175(A):324–335. <https://doi.org/10.1016/j.neucom.2015.10.065>
18. Balazia M, Plataniotis KN (2016) Human gait recognition from motion capture data in signature poses. *IET Biom* 6(2):129–137. <https://doi.org/10.1049/iet-bmt.2015.0072>
19. Rida I, Almaadeed S, Bouridane A (2016) Gait recognition based on modified phase-only correlation. *Signal Image Video Process* 10(3):463–470
20. Alotaibi M, Mahmood A (2017) Improved gait recognition based on specialized deep convolutional neural network. *Comput Vis Image Underst* 164:103–110
21. Wu Z, Huang Y, Wang L, Wang X, Tan T (2016) A comprehensive study on cross-view gait based human identification with deep CNNs. *IEEE Trans Pattern Anal Mach Intell* 39(2):209–226
22. Carley C, Ristani E, Tomasi C (2019) Person re-identification from gait using an autocorrelation network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*

23. Muramatsu D, Makihara Y, Yagi Y (2015) View transformation model incorporating quality measures for cross-view gait recognition. *IEEE Trans Cybern* 46(7):1602–1615
24. Zheng S, Zhang J, Huang K, He R, Tan T (2011) Robust view transformation model for gait recognition. In: *International Conference on Image Processing (ICIP)*. Belgium, Brussels
25. Liu G, Zhong S, Li T (2019) Gait recognition method of temporal-spatial HOG features in critical separation of Fourier correction points. *Future Generat Comput Syst* 94:11–15. <https://doi.org/10.1016/j.future.2018.09.012>
26. Connie T, Goh MKO, Teoh ABJ (2018) Human gait recognition using localized Grassmann mean representatives with partial least squares regression. *Multimed Tools Appl* 77(21):28457–28482
27. Hu B, Gao Y, Guan Y, Long Y, Lane N, Ploetz T (2018) Robust cross-view gait identification with evidence: a discriminant gait GAN (DiGGAN) approach on 10000 people. *arXiv preprint arXiv:1811.10493*
28. Chao H, He Y, Zhang J, Feng J (2019) Gaitset: regarding gait as a set for cross-view gait recognition. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 33:8126–8133
29. Wang X, Feng S, Yan WQ (2019) Human gait recognition based on self-adaptive hidden Markov model. In: *IEEE/ACM Transactions on Computational Biology and Bioinformatics*
30. Li S, Liu W, Ma H (2019) Attentive spatial-temporal summary networks for feature learning in irregular gait recognition. In: *IEEE Transactions on Multimedia*
31. Zhang Z, Tran L, Liu F, Liu X (2019) On learning disentangled representations for gait recognition. *arXiv preprint arXiv:1909.03051*
32. He Y, Zhang J, Shan H, Wang L (2018) Multi-task gans for view-specific feature learning in gait recognition. *IEEE Trans Inf Forensics Secur* 14(1):102–113
33. Liao R, Yu S, An W, Huang Y (2020) A model-based gait recognition method with body pose and human prior knowledge. *Pattern Recognit* 98:107069
34. Bobick AF, Davis JW (2001) The recognition of human movement using temporal templates BT-pattern analysis and machine intelligence. *IEEE Trans Pattern Anal Mach Intell* 23(3):257–267
35. Han J, Bhanu B (2006) Individual recognition using gait energy image. *IEEE Trans Pattern Anal Mach Intell* 28(2):316–322. <https://doi.org/10.1109/TPAMI.2006.38>
36. Ben X, Zhang P, Lai Z, Yan R, Zhai X, Meng W (2019) A general tensor representation framework for cross-view gait recognition. *Pattern Recognit* 90:87–98
37. Wang Y, Song C, Huang Y, Wang Z, Wang L (2019) Learning view invariant gait features with two-stream GAN. *Neurocomputing* 339:245–254
38. Takemura N, Makihara Y, Muramatsu D, Echigo T, Yagi Y (2018) Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSI Trans Comput Vis Appl* 10(4):1–14
39. Iwama H, Okumura M, Makihara Y, Yagi Y (2012) The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition. *IEEE Trans Inf Forensics Secur* 7(5):1511–1521
40. Elharrouss O, Al-Maadeed N, Al-Maadeed S (2019) Video summarization based on motion detection for surveillance systems. In: *2019 IEEE 15th International Wireless Communications and Mobile Computing Conference (IWCMC)*, 366–371
41. Moujahid D, Elharrouss O, Tairi H (2018) Visual object tracking via the local soft cosine similarity. *Pattern Recognit Lett* 110:79–85
42. He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: *Proceedings of the IEEE International Conference on Computer Vision*, 1026–1034
43. Bloisi DD, Pennisi A, Iocchi, L (2017) Parallel multi-model background modeling. *Pattern Recognit Lett* 96:45–54
44. Elharrouss O, Abbad A, Moujahid D, Tairi H (2017) Moving object detection zone using a block-based background model. *IET Comput Vis* 12(1):86–94

Affiliations

**Omar Elharrouss¹ · Noor Almaadeed¹ · Somaya Al-Maadeed¹ ·
Ahmed Bouridane²**

Noor Almaadeed
n.alali@qu.edu.qa

Somaya Al-Maadeed
s_alali@qu.edu.qa

Ahmed Bouridane
ahmed.bouridane@northumbria.ac.uk

¹ Department of Computer Science and Engineering, Qatar University, Doha, Qatar

² Department of Computer Science and Digital Technologies, Northumbria University, Newcastle Upon Tyne, UK